

University of Groningen

Instruments used to assess functional limitations in workers applying for disability benefit

Spanjer, Jerry; Groothoff, Johan W.; Brouwer, Sandra

Published in:
Disability and Rehabilitation

DOI:
[10.3109/09638288.2011.570413](https://doi.org/10.3109/09638288.2011.570413)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Early version, also known as pre-print

Publication date:
2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Spanjer, J., Groothoff, J. W., & Brouwer, S. (2011). Instruments used to assess functional limitations in workers applying for disability benefit: a systematic review. *Disability and Rehabilitation*, 33(23-24), 2143-2150. <https://doi.org/10.3109/09638288.2011.570413>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

REVIEW

Instruments used to assess functional limitations in workers applying for disability benefit: a systematic review

JERRY SPANJER^{1,2}, JOHAN W. GROOTHOFF² & SANDRA BROUWER²

¹UWV, Employee Benefits Insurance Authority, Amsterdam, The Netherlands and ²Department of Health Sciences, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Accepted March 2011

Abstract

Purpose. To systematically review the quality of the psychometric properties of instruments for assessing functional limitations in workers applying for disability benefit.

Method. Electronic searches of Medline, Embase, CINAHL and PsycINFO were performed to identify studies focusing on the psychometric properties of instruments used to assess functional limitations in workers' compensation claimants. Two independent reviewers applied the inclusion criteria to select relevant articles and then evaluated the psychometric qualities of the instruments found.

Results. Of the 712 articles that were identified, 10 studies met the inclusion criteria, reporting on four instruments: the Roland–Morris Disability Questionnaire (RMDQ), the Patient-Specific Functional Scale (PSFS), the Isernhagen Work System (IWS) and the Multiperspective Multidimensional Pain Assessment Protocol (MMPAP). The questionnaires (RMDQ and PSFS) did not focus specifically on the work situation and measured three to eight functional limitations. The psychometric qualities of the IWS were poor to moderate. For the MMPAP, only predictive validity was measured. The instruments assessed a range varying between 3 and 34 physical functional limitations. No instruments were found for assessing mental limitations in workers' compensation claimants.

Conclusion. Studies on four instruments specifically focusing on assessing physical functional limitations in workers applying for disability benefit were found. All four instruments have limitations regarding their psychometric qualities or contents. Since the RMDQ has the best demonstrated psychometric qualities and takes little time to complete it, we recommend the RMDQ for clinicians in rehabilitation. For the assessment of functional limitations in workers applying for disability benefit a combination of questionnaires, performance tests or interviews together with the judgment by physicians looks the most promising.

Keywords: Systematic review, assessment, work capacity evaluation, workers' compensation, instruments, psychometrics, disability benefit

Introduction

If workers become disabled in The Netherlands, their employers still have to pay wages for 2 years. Then a workers' compensation benefit from the Social Security Office can be applied for. The benefit is based on the loss of the wage-earning capacity of the claimant. An insurance physician first assesses the functional limitations, then a labour expert assesses what the claimant theoretically is still able to earn in suitable work. The insurance physician bases the assessment of the functional limitations on an interview with the claimant, a medical examina-

tion and information from treating physicians; and usually uses no other specific instruments. The reliability and validity of these assessments are questionable [1,2].

The assessment of functional limitations is an important part of the evaluation for the work disability pension and has immense individual, financial and social consequences. According to the International Classification of Functioning, Disability and Health (ICF), limitations in functioning are defined as the limitations in performing a task or action by an individual [3]. Because this is an umbrella term encompassing all body functions,

activities and participation [3], in the present study, the definition of functional limitations is: limitations in or inability to perform certain physical activities such as walking and lifting, or mental activities such as concentrating and conflict handling. Therefore, functional limitations can be distinguished from symptoms (such as pain and fatigue), activity limitations (such as self-care tasks and gardening) and participation restrictions (such as leisure time activities and work).

Instruments used to assess functional limitations have to provide reliable and valid information to enable appropriate decisions about the work disability pension [4,5]. Reliability is the extent to which an instrument is free from error and consistent over time, between different raters or between parts of the test [4]. Validity is considered to be the extent to which an instrument measures what it is intended to measure [5].

Several instruments to measure or assess functional limitations are described in literature [6,7], often with satisfactory levels of reliability and validity. Among them are questionnaires like the Oswestry Disability Index, the Pain Disability Index, the Roland-Morris Disability Questionnaire and the Upper Extremity Functional Scale [8–10]; and functional capacity evaluations or performance tests [11,12]. However there are some concerns as to their relevance and validity in assessing functional limitations in workers applying for disability benefit:

- often the reliability and validity of these instruments were measured in rehabilitation patients or in job fitness for healthy people and not in workers applying for disability benefit. Because an instrument is not validated *per se*, but rather its use in a specific setting or in a specific target group [13,14], an instrument can be valid in a rehabilitation setting but not in the assessment of workers applying for disability benefit. For instance, if workers believe a certain test could affect their ability to receive benefits, this may affect outcomes [15,16]. Therefore, the validity of these instruments in workers applying for disability benefit is questionable.
- questionnaires often do not have a work-related point of reference: they consider activities, such as self-care tasks, domestic tasks and sickness absence; but whether the patient can lift 10 kg at work remains unknown [8].
- the items of the instruments often inquire about a combination of symptoms, functional limitations and about activity and participation restrictions, rather than focusing only on functional limitations [17].

Until now, no reviews have been published concerning instruments that assess mental and/or physical functional limitations in workers applying for disability benefit. To fill this gap, we systematically reviewed the literature on instruments for assessing functional limitations in workers applying for disability benefit. For those instruments found, we evaluated the psychometric properties, such as reliability and validity.

Methods

Literature search

Studies were identified by searches of the electronic bibliographic databases Medline (biomedical literature), Embase (biomedical and pharmacological literature), CINAHL (nursing and allied health literature) and PsycINFO (psychological literature). The searches were limited to literature published between 1980 and December 2008. The search terms used were: Disability evaluation [Major topic]; AND Observer variation [Mesh] OR Psychometrics [Mesh] OR Reproducibility of Results [Mesh] OR reliability OR validity; AND Workers' Compensation [Mesh] OR Work* OR claimants OR job* OR occupation* OR vocation*.

Articles were included in the review if all of the following criteria were met:

1. An instrument was described for assessing functional limitations in a work setting.
2. The instrument was used among workers applying for disability benefit.
3. The article was published in English, German or Dutch.
4. The article was a primary peer-reviewed research study.
5. Psychometric properties of the instrument were presented.

Excluded were studies regarding return to work without assessing functional limitations, a clinical or rehabilitation setting without assessing the eligibility for a workers' compensation benefit, job fitness in healthy people and malingering. We also excluded case studies, letters to the editor and book chapters. Review papers were only used to screen for further original papers. References of retrieved articles were screened for additional relevant studies.

Applying these criteria, two of the authors (JS, insurance physician and SB, researcher with experience in work and health research) independently reviewed the titles and abstracts of the literature to identify potentially relevant articles. If the title and abstract did not provide enough information to

decide whether or not the inclusion criteria were met, the article was included for full-text selection. From the articles included, we read the full text and the same two reviewers applied the inclusion criteria to the full text. Disagreements between the reviewers were discussed and resolved during a consensus meeting.

The names of identified instruments were used as terms for a further search of the electronic databases. We systematically reviewed the literature on their reliability and validity. To identify eligible studies we used the above-mentioned electronic bibliographic databases with the following keywords: 'the name of the instrument'; AND Psychometrics [Mesh] OR Reproducibility of Results [Mesh] OR reliability OR validity.

Quality assessment

The psychometric properties of the instruments were assessed by the two authors (JS and SB) independently. Definitions, analysis and interpretation of content validity, internal consistency, criterion validity, construct validity, reproducibility (agreement and reliability), responsiveness, and floor and ceiling effects were rated using the criteria described by Terwee et al. [18] and Innes and Straker [4,5] (Table I). Possible ratings for the psychometric properties were good (+), moderate (\pm) and poor (−).

Description of the instruments

Descriptive data for the instruments found were extracted from the publications, and included the type of instrument (questionnaire, performance test, physician assessment), time to complete, number of items, number of scales and the target population for which the instrument had been developed.

Results

The search in Medline yielded 439 articles, CINAHL 174, PsycINFO 17 and Embase 281. From this total of 911 articles, 712 remained after removal of duplicates. Reference tracing resulted in one additional article. A total of 676 articles were read by title and abstract only, 37 articles were selected for full text reading. The most important reason for excluding studies was because they were not conducted in workers applying for a disability benefit but in a clinical or rehabilitation setting without assessing the eligibility for a workers' compensation benefit or did not measure functional limitations. Fifteen articles were not published in English, German or Dutch. A total of ten studies fulfilled the inclusion criteria, reporting on four different instruments: the Roland-Morris Disability Questionnaire (RMDQ) [19,20], the Patient-Specific Functional Scale (PSFS) [21], the Isernhagen Work

Table I. Quality criteria used for psychometric properties [4,18].

Property	Definition	Quality criteria
1. Content validity	The extent to which the domain of interest is comprehensively sampled by the items in the test	Poor: the test does not measure what it is intended to measure Moderate: not all relevant components are included Good: the test measures all relevant components
2. Internal consistency	The extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct	Cronbach's $\alpha \leq 0.70$ poor; 0.71–0.80 moderate; > 0.80 good
3. Criterion validity	The extent to which scores on a particular test relate to another valued measure	Correlation between test and the criterion measure $r \leq 0.50$ poor; 0.51–0.75 moderate; > 0.75 good
4. Construct validity	The extent to which a test is well correlated with a hypothetical construct or theoretical expectation	$> 75\%$ of the results are in accordance with the theoretical expectation. Convergence between tests: poor $r \leq 0.30$; moderate 0.31–0.60; good > 0.60
5. Reproducibility		
5.1. Agreement	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	MIC $<$ SDC OR MIC outside the LOA OR convincing arguments that agreement is acceptable
5.2. Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error)	ICC: poor < 0.75 ; moderate 0.75–0.90; good > 0.90 Kappa: poor < 0.40 ; moderate 0.40–0.60; good > 0.60
6. Responsiveness	The ability of a test to detect clinically important changes over time	Significant difference in t -test: $p < 0.05$ Effect Size and Standard Response Mean: poor: 0.20–0.50; moderate 0.51–0.80; large > 0.80
7. Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score	$\leq 15\%$ of the respondents achieved the highest or lowest possible scores

MIC = minimal important change; SDC = smallest detectable change; LOA = limits of agreement; ICC = intraclass correlation.

System (IWS) [22–27] and the Multiperspective Multidimensional Pain Assessment Protocol (MMPAP) [28].

An overview of these studies and the psychometric properties of the instruments are described in Table II. For all instruments, the predictive validity was described; for all instruments but the MMPAP, construct validity was studied. The RMDQ was the only instrument for which responsiveness, internal consistency and floor and ceiling effects were measured. None of the studies examined the face and content validity or reproducibility of the instruments in workers applying for disability benefit.

Below we will describe the content of the four instruments and their psychometric properties investigated in workers applying for disability benefit (Table III).

Roland–Morris Disability Questionnaire

The RMDQ [29] is a questionnaire that is widely used to assess self-reported physical disability associated with low back pain. The 24 dichotomous items cover symptoms (4 items, e.g. pain and poor appetite), activity limitations and participation restrictions (12 items, e.g. staying in bed and avoiding heavy work) and functional limitations (8 items, e.g. walking, standing and climbing stairs). Total scores range from 0 (no disability) to 24 (severe disability). The time to administer the questionnaire is 5–10 min. In workers applying for disability benefit, the RMDQ showed good internal consistency (Cronbach's α 0.95), no floor and ceiling defects, and moderate to large responsiveness (response mean 0.78–0.84 for improvement). Construct validity was good with positive correlations with the

Table II. Studies describing instruments for assessing functional limitations in workers applying for disability benefit.

Instrument	Population	Psychometrics	Outcome	1st author, year
Roland–Morris Disability Questionnaire (RMDQ)	N = 284 D = back pain T = median 56 days	Internal consistency	+	Turner, 2003 [19]
		Floor and ceiling effects	+	
		Responsiveness	\pm /+	
		Construct validity (SF-12, SF-36)	+	
Patient-Specific Functional Scale (PSFS)	N = 959 D = back pain T = 0–90 days	Predictive validity (return to work)	+	Baldwin, 2007 [20]
		Construct validity (SF-12, NRS-101)	+	
		Predictive validity (time receiving benefit; time to claim closure; recurrence)	\pm / \pm /–	
		Construct validity (PDI, SF-36, VAS)	\pm / \pm /–	
Isernhagen Work System (IWS)	N = 321 D = musculoskeletal disorder T = mean 388 days	Predictive validity (time receiving benefit; time to claim closure; recurrence)	\pm /–/–	Gross, 2003, 2004 (2 \times), 2005, 2006 [22–26]
		Construct validity (PDI, VAS)	\pm	
		Predictive validity (time receiving benefit; time to claim closure; recurrence)	\pm / \pm /–	
		Construct validity (PDI, VAS)	\pm	
Multiperspective Multidimensional Pain Assessment Protocol (MMPAP)	N = 336 D = upper extremity disorder T = mean 468 days	Predictive validity (time receiving benefit; time to claim closure; recurrence)	\pm / \pm /–	Gross, 2006 [27]
		Predictive validity (employment after 8 months)	+	
		Predictive validity (employment after 8 months)	+	
		Predictive validity (employment after 8 months)	+	
Multiperspective Multidimensional Pain Assessment Protocol (MMPAP)	N = 599 D = chronic pain T > 180 days	Predictive validity (employment after 8 months)	+	Rucker, 1995 [28]
		Predictive validity (employment after 8 months)	+	
		Predictive validity (employment after 8 months)	+	
		Predictive validity (employment after 8 months)	+	

N = number of patients; D = disorder; T = days from injury to test.

Outcome: minus; = poor; \pm = moderate; + = good.

NRS = Numeric Rating Scale; PDI = Pain Disability Index; SF = Short-Form; VAS = Visual Analogue Scale for pain.

Table III. Description of the instruments.

	Way of assessment	Time to complete	Number of items	Item scale	Target population
RMDQ	Questionnaire	5–10 min	24	Dichotomous (yes/no)	low back pain
PSFS	Questionnaire	5 min	3–5	10 point	musculoskeletal problems
IWS	Performance test	5 h in 2 days	28	Numeric, ratio	musculoskeletal problems
MMPAP	2 Physicians + Questionnaire	2–4 h	65	3–6 point	chronic pain

RMDQ, Roland–Morris Disability Questionnaire; PSFS, Patient-Specific Functional Scale; IWS, Isernhagen Work System; MMPAP, Multiperspective Multidimensional Pain Assessment Protocol.

Numeric Rating Scale (NRS-101) and the Short Form-12 and -36 (SF-12 and SF-36) ($r=0.70-0.85$). Predictive validity was good with positive prediction of return to work outcomes [19,20].

Patient-Specific Functional Scale

The PSFS [30] is a patient-specific outcome measure, which was designed for use in patients with varied musculoskeletal problems. It measures functional status by asking the patient to name up to five activities which they are having difficulty performing because of their problem. After specifying the activities, patients are asked to rate each activity on a 0–10 scale, with 0 representing the inability to perform the activity and 10 the ability to perform as well as before the onset of symptoms. Time to administer is about 5 min. In workers applying for disability benefit the PSFS showed moderate construct validity with correlations of 0.32–0.53 for the Pain Disability Index (PDI), 0.32–0.44 for the SF-36 and 0.19–0.24 for the Visual Analogue Scale for pain (VAS). As for predictive validity, the association between functional scales and suspension of benefits, an adjusted hazard ratio (HR) of 1.16 (95% CI, 1.07–1.27) was found, and for claim closure an HR of 1.14 (95% CI, 1.06–1.22). No significant association with recurrence of claiming was found [21].

Isernhagen Work System

Functional capacity evaluations (FCE) are standardised tests which measure the actual physical performance of the patient or worker, such as lifting, carrying, pushing and walking. There are several FCE protocols, and the IWS is one of them. The IWS consists of 28 tests that reflect work-related activities. During administration of the IWS, the clinician relies on observation of biomechanical and physiological signs of effort to determine safe, maximal performance levels [31]. It is a 2-day test, taking 2–3 h each day, with some tests being repeated on day 2. Patients' performance on the test can be matched to the specific tasks of jobs. In workers applying for disability benefit, the IWS showed moderate construct validity in correlation with the PDI ($r=0.44-0.55$) and the VAS ($r=0.34-0.45$). As for predictive validity, better lift performance was associated with a shorter time receiving benefit (for upper extremity disorder HR 1.55; 95% CI 1.29–1.87; for low back pain HR 1.48; 95% CI 1.14–1.92) and time to claim closure (for upper extremity disorder HR 1.81; 95% CI 1.49–2.20; for low back pain HR 1.17; 95% CI 0.91–1.50) [26–31].

There was no association [27] or even a negative association [24] with recurrence.

Multiperspective Multidimensional Pain Assessment Protocol

The MMPAP is a pain assessment tool that collects and uses information from patient self-reports and medical examination, along with an assessment by two physicians separately. Major domains assessed in the protocol include pain, mental health status, social support, medical information, functional limitations and abilities and rehabilitation potential [28]. The patient's assessment of the functional limitations includes 17 items in the activities of daily living (ADL) domain (e.g. climbing stairs and travelling); the physician's rating of the functional abilities domain also includes 17 items (e.g. standing, sitting and lifting). The length of time to complete the MMPAP was patient-specific and varied from 2 to 4 h. In workers applying for disability benefit, the MMPAP showed a good predictive validity (90% for outcome employment) [28].

Discussion

We systematically reviewed the literature on instruments that assess functional limitations in workers applying for disability benefit and found studies on four instruments: two questionnaires (RMDQ, PSFS), a performance test (IWS) and an instrument which combined a questionnaire and examination by physicians (MMPAP). In workers applying for disability benefit for all of these four instruments, the predictive validity was assessed, for three of them the construct validity, and for none of them the reliability. The predictive validity was good in the RMDQ and MMPAP, and ranged from poor to moderate in the IWS and PSFS. The construct validity was good in the RMDQ, poor to moderate in the IWS and PSFS, and not measured in the MMPAP. The instruments assessed a range varying between 3 and 34 physical functional limitations. No instruments were found for assessing mental functional limitations in workers applying for disability benefit, even though psychological complaints are responsible for 35% of the claims for workers' compensation in The Netherlands [32].

This is the first review focusing on instruments for assessing functional limitations in workers applying for disability benefit. Because we reviewed the literature systematically in four major electronic databases and checked for additional literature in references, we assume we have included all relevant instruments to assess functional limitations in workers applying for disability benefit. We used the

MeSH term 'disability evaluation' as search term, which is defined as: 'Determination of the degree of a physical, mental, or emotional handicap. The diagnosis is applied to legal qualification for benefits and income under disability insurance and to eligibility for Social Security and workmen's compensation benefits', which is covering the subject we wanted to study. This includes the MeSH term 'work capacity evaluation' and is a broad search term, having about 12,000 hits in Medline. Therefore, we think the search has not been too restrictive and cannot be an explanation for finding only four instruments.

In clinical and rehabilitation settings, the psychometric properties of the four instruments we found have been investigated more frequently. In a rehabilitation setting, the RMDQ [8,33] and the PSFS [30,34–36] have been demonstrated as valid, reliable and responsive to change. As for the IWS, in a rehabilitation setting, inter-rater reliability and predictive validity were good; concurrent validity was low to moderate [12]. The MMPAP proved to be a reliable and valid tool in a population of 651 patients [37]. Our findings on the psychometric qualities of RMDQ and IWS in claimants were in line with these studies, although in most of the studies the instruments were used to evaluate the results of therapy in a clinical setting and not as an instrument to assess functional limitations in workers applying for a disability benefit.

Assessing the validity of instruments measuring functional limitations is a problem because there is no gold standard, and often it is unclear whether it is performance or capacity that has to be assessed. For instance, does a patient with non-specific low back pain have a reduced lifting capacity? And if so, how many kilograms is the patient able to lift? In the studies identified in this review the predictive validity was assessed by measuring 'return to work' or the time during which the claimant did receive a benefit. However, these measures may underestimate functional limitations, for instance, when a patient with functional limitations changes functional job status, resumes work part-time or cannot find a job [38]. They may also overestimate the patient's functional limitations, for instance, if a patient is not motivated to go to work. Therefore, 'return to work' and 'time receiving benefit' are not well suited for measuring validity in instruments that measure functional limitations.

We found two questionnaires (RMDQ and PSFS) that measured functional limitations in workers applying for disability benefit. Looking at the content of both questionnaires, there are some questions as to their suitability for assessing functional limitations in workers applying for disability benefit. The questionnaires are not work oriented, but mainly measure limitations in daily life, which makes them less suitable

for work rehabilitation. There is a mixed content: not only functional limitations are measured but also activity limitations and participation restrictions. Only a few functional limitations are assessed and in the RMDQ there is no grading of the functional limitations. Furthermore, there is only a registration of the perceived limitations, but no real assessment is being conducted. For instance, if a patient claims he cannot walk at all, no assessment is made if this is reasonable considering the patient's disease.

The IWS, on the other hand, is work oriented, measures 28 physical items and gives a grading of these items. It can be used for all somatic disorders instead of a specific disorder such as low back pain. Unfortunately, predictive as well as construct validity in the studies we found was poor to moderate and no studies on reliability for workers applying for a disability benefit could be found. The IWS measures patients' performance; in addition, there has to be an assessment of the sincerity of the patient's effort, the ability to perform work outside a laboratory setting, and whether activities are considered medically safe [39]. One disadvantage is the fact that it takes 2 days and 2–3 h each day to execute the IWS. An abbreviated IWS which measures only three items (lifting, crouching and standing) and takes only 1 h to execute was found to predict future work status in low back disorders comparably to the entire IWS. Therefore, an abbreviated IWS may offer an efficient alternative [26].

Questionnaires and performance tests are designed to assess patients' actual functioning or performance in a given situation. In the assessment of functional limitations in worker applying for disability benefit, however, the capacity to perform is the issue that has to be assessed. For instance, if a patient does not perform a certain task, this can be due to a motivational problem or inadequate behaviour. In daily practice, physicians often play an important role in this assessment [40]. Therefore, it is remarkable that almost no studies are conducted into the reliability and validity of physicians' assessments. We only found one study in the year 1995 [28] that described a protocol for physicians, the MMPAP. Although research into claimants was limited to predictive validity, in a clinical setting satisfactory reliability and validity was found.

Recommendations

Although self-report questionnaires cannot be the only measure to base a work disability benefit upon, they can serve as a guide and inventory taking. Depending on the patient's complaints different questionnaires can be used like the RMDQ in low back pain patients or the Upper Extremity Functional

Scale in patients with shoulder complaints. If necessary, additional performance tests can be used, preferably abbreviated tests depending on the diagnosis. Questionnaires and performance tests alone cannot properly assess functional limitations without an appraisal of the outcome of these tests. Therefore, physicians are needed to weight the outcome of these tests against medical information like physical examination, X-rays, laboratory findings and medical knowledge. A combination of patient self-reports, performance tests and medical examination with an assessment by physicians, like the MMPAP, seems the most solid tool for assessing functional limitations in work disability benefits.

It is important to develop specific questionnaires for specific goals. The RMDQ, e.g. can well be used to measure the effect of therapy and the consequences of low back pain in daily life. For work reintegration or assessing the functional limitations in work, disability assessment more work oriented items are required. Semi-structured interview protocols [41] can also give specific information that can well be used in work disability assessment. Because an appraisal of the acquired information (from questionnaires, performance tests or interviews) is always needed, psychometric properties can best be measured after the appraisal phase.

To summarise, in the present review study, four instruments were found specially focusing on assessing the functional limitations in workers applying for disability benefit. Of these four instruments found, only a limited amount of functional limitations were measured, or the psychometric qualities were not satisfactorily demonstrated in work disability assessments. Since the RMDQ has the best demonstrated psychometric qualities and takes little time to complete it, we recommend the RMDQ for clinicians in rehabilitation. Because the assessment of functional limitations in workers applying for disability benefit not only has immense individual implications, but also implications for society as a whole, more evidence-based instruments need to be developed in future research. Performance tests and questionnaires alone cannot properly assess functional limitations without an appraisal of the outcome of these tests. A combination of gathering information with instruments such as questionnaires, performance tests or interviews together with interpretation and judgment by physicians looks the most promising.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- Spanjer J. The reproducibility in disability assessment: a literature search. *Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde* 2001;9:195–198.
- Söderberg E, Alexanderson K. Sickness certification practices of physicians: a review of the literature. *Scand J Public Health* 2003;31:460–474.
- WHO. International classification of functioning, disability and health. Geneva: World Health Organisation; 2001.
- Innes E, Straker L. Reliability of work-related assessments. *Work* 1999;13:107–124.
- Innes E, Straker L. Validity of work-related assessments. *Work* 1999;13:125–152.
- Matheson LN, Kaskutas V, McCowan S, Shaw H, Webb C. Development of a database of functional assessment measures related to work disability. *J Occup Rehabil* 2001;11:177–199.
- Flores L, Gatchel RJ, Polatin PB. Objectification of functional improvement after nonoperative care. *Spine* 1997;22:1622–1633.
- Grotle M, Brox JI, Vollestad NK. Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. *Spine* 2005;30:130–140.
- Williams RM, Schmuck G, Allwood S, Sanchez M, Shea R, Wark G. Psychometric evaluation of health-related work outcome measures for musculoskeletal disorders: a systematic review. *J Occup Rehabil* 2007;17:504–521.
- Wind H, Gouttebarger V, Kuijer PP, Frings-Dresen MHW. Assessment of functional capacity of the musculoskeletal system in the context of work, daily living, and sport: a systematic review. *J Occup Rehabil* 2005;15:253–272.
- Innes E. Reliability and validity of functional capacity evaluations: an update. *Int J Disabil Manage Res* 2006;1:135–148.
- Gouttebarger V, Wind H, Kuijer PP, Frings-Dresen MH. Reliability and validity of Functional Capacity Evaluation methods: a systematic review with reference to Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen work system. *Int Arch Occup Environ Health* 2004;77:527–537.
- Schultz IZ, Crook J, Berkowitz J, Milner R, Meloche GR. Predicting return to work after low back injury using the Psychosocial Risk for Occupational Disability Instrument: a validation study. *J Occup Rehabil* 2005;15:365–376.
- Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 2nd ed. Upper Saddle River, New Jersey: Prentice Hall, Inc.; 2000.
- Menard MR, Hoens AM. Objective evaluation of functional capacity: medical, occupational, and legal settings. *J Orthop Sports Phys Ther* 1994;19:249–260.
- Kreider B. Latent work disability and reporting bias. *J Hum Resour* 1999;34:734–769.
- Mayer J, Mooney V, Matheson L, Leggett S, Verna J, Balourdas G, DeFilippo G. Reliability and validity of a new computer-administered pictorial activity and task sort. *J Occup Rehabil* 2005;15:203–213.
- Terwee CB, Bot SDM, de Boer MR, van der Windt DLAWM, Knol DL, Dekker J, Bouter LA, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- Turner JA, Fulton-Kehoe D, Franklin G, Wickizer TM, Wu R. Comparison of the Roland-Morris Disability Questionnaire and generic health status measures: a population-based study of workers' compensation back injury claimants. *Spine* 2003;28:1061–1067.
- Baldwin ML, Butler RJ, Johnson WG, Cote P. Self-reported severity measures as predictors of return-to-work outcomes in occupational back pain. *J Occup Rehabil* 2007;17:683–700.

21. Gross DP, Battie MC, Asante AK. The Patient-Specific Functional Scale: validity in workers' compensation claimants. *Arch Phys Med Rehabil* 2008;89:1294–1299.
22. Gross DP, Battie MC. Construct validity of a kinesiophysical functional capacity evaluation administered within a worker's compensation environment. *J Occup Rehabil* 2003;13:287–295.
23. Gross DP, Battie MC, Cassidy JD. The prognostic value of functional capacity evaluation in patients with chronic low back pain. Part 1: timely return to work. *Spine* 2004;29:914–919.
24. Gross DP, Battie MC. The prognostic value of functional capacity evaluation in patients with chronic low back pain. Part 2: sustained recovery. *Spine* 2004;29:920–924.
25. Gross DP, Battie MC. Factors influencing results of functional capacity evaluations in workers' compensation claimants with low back pain. *Phys Ther* 2005;85:315–322.
26. Gross DP, Battie MC, Asante A. Development and validation of a short-form functional capacity evaluation for use in claimants with low back disorders. *J Occup Rehabil* 2006;16:53–62.
27. Gross DP, Battie MC. Does functional capacity evaluation predict recovery in workers' compensation claimants with upper extremity disorders? *Occup Environ Med* 2006;63:404–410.
28. Rucker KS, Metzler HM. Predicting subsequent employment status of SSA disability applicants with chronic pain. *Clin J Pain* 1995;11:22–35.
29. Roland M, Morris R. A study of the natural history of back pain. I. Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983;8:141–144.
30. Stratford P, Gill C, Westaway M, Binkley J. Assessing disability and change on individual patients: a report of a patient specific measure. *Physiother Can* 1995;47:258–263.
31. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgments of level of lift effort in a kinesiophysical Functional Capacity Evaluation. *Work* 1999;12:145.
32. Statline, Dutch Bureau for Statistics [Internet]. The Hague, Netherlands. Electronic Citation. <http://statline.cbs.nl>. Last accessed 12 February 2010.
33. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 2002;82:8–24.
34. Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther* 1998;27:331–338.
35. Cleland JA, Fritz JM, Whitman JM, Palmer JA. The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. *Spine* 2006;31:598–602.
36. Jolles BM, Buchbinder R, Beaton DE. A study compared nine patient-specific indices for musculoskeletal disorders. *J Clin Epidemiol* 2005;58:791–801.
37. Rucker KS, Metzler HM, Kregel J. Standardization of chronic pain assessment: a multiperspective approach. *Clin J Pain* 1996;12:94–110.
38. Fishbain DA, Rosomoff HL, Goldberg M, Cutler R, Abdel-Moty E, Khalil TM, Rosomoff RS. The prediction of return to the workplace after Multidisciplinary Pain Center treatment. *Clin J Pain* 1993;9:3–15.
39. Pransky GS, Dempsey PG. Practical aspects of functional capacity evaluations. *J Occup Rehabil* 2004;14:217–229.
40. De Boer WEL, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. *Pratiques et Organisations de Soins* 2007;38:205–217.
41. Spanjer J, Krol B, Brouwer S, Popping R, Groothoff JW, Klink JJL van der. Reliability and validity of the Disability Assessment Structured Interview (DASI): a tool for assessing functional limitations in claimants. *J Occup Rehabil* 2010;20:33–40.